PROGETTO LAUREE SCIENTIFICHE
SETTIMANA MATEMATICA

Laboratorio n.2:
La probabilità applicata alle scienze sociali: il teorema del ballottaggio.

Docente responsabile: Rita Giuliano
Collaboratori: Andrea Carpignani e Annalisa Castellucci
Tutor: Pietro Battiston e Giulio Tiozzo

Il laboratorio si proponeva di esporre agli studenti un teorema di probabilità interessante sia per la sua aderenza alla realtà che per la sua formulazione particolarmente semplice che nasconde però un concetto decisamente non ovvio.

Siano A e B due candidati ad un'elezione (da qui il nome del teorema), che ricevono rispettivamente a e b voti, con $a>b$. Supponendo che lo scrutinio avvenga un voto alla volta, la probabilità che, nel conteggio parziale dei voti, il candidato A sia sempre in vantaggio è:

$$\frac{(a-b)}{(a+b)}$$

Se chiamiamo rispettivamente α e β le percentuali di voti ottenuti da A e B , con $\alpha+\beta=100\%$ (ovvero scegliendo di trascurare eventuali voti non validi) e ovviamente $\alpha>\beta$, la stessa probabilità è esprimibile come $\alpha-\beta$.

Il teorema è stato dimostrato nel 1924 da A. Aeppli, e ne sono state formulate anche numerose generalizzazioni, che vertono principalmente sull'aumento del numero dei candidati o su modi alternativi di rappresentare i rispettivi numeri di voti.

L'organizzazione del lavoro:

Il laboratorio si è articolato in 4 fasi:

1. la presentazione del teorema
2. l'esposizione, con la lettura individuale delle dispense distribuite agli studenti e successiva discussione, di alcune nozioni base di probabilità e calcolo combinatorio
3. la dimostrazione del teorema applicando il *principio di riflessione*
4. la sperimentazione, al computer

Prima fase: la presentazione del teorema.

Sebbene una dimostrazione completa del teorema sia stata trovata solo nel secolo scorso, alcuni casi specifici erano stati trattati già da tempo. Ad esempio, è del 1527 una trattazione del caso particolare in cui $a+b\leqslant 10$, da parte di Apianus, mentre successivi studi furono compiuti, per valori di a e b sempre maggiori, da Cardano, Mersenne, e Pascal. Invece DeMoivre, studiando nel 1708 il problema della *rovina del giocatore*, fornì una dimostrazione che, sebbene lui non l'abbia collocata in questo contesto, dimostra il teorema del ballottaggio nel caso particolare in cui $a=b$.

Si è spiegato ai ragazzi che la scelta del teorema non è stata casuale, ma motivata dalla sua attinenza alla realtà (forse nel contesto delle ultime elezioni politiche in Italia, caratterizzate da un margine minimo tra i due schieramento e da proiezioni che davano con il passare delle ore risultati altalenanti, ad alcuni ragazzi è tornato in mente il problema, nel suo caso particolare $a-b\approx 0$ ...), e dal fatto che troppo spesso lo studio delle probabilità venga accostato, in ambito divulgativo, soltanto a problemi di gioco d'azzardo, dandone un'immagine distorta e limitata.

Seconda fase: elementi di probabilità e calcolo combinatorio.

E' risultato evidente che lo studio delle probabilità, più del calcolo combinatorio, presenta alcuni concetti che l'intuizione dei ragazzi arriva a comprendere senza difficoltà. Ad esempio nessuno aveva dubbi riguardo al fatto che la probabilità dell'unione di due eventi disgiunti sia la somma delle probabilità dei due eventi. Va detto che gran parte dei ragazzi provenivano da classi quarte e quinte di licei scientifici, e la maggior parte di loro aveva già affrontato un po' di studio di probabilità a scuola. Ma anche per gli altri non c'è stato bisogno di introdurre in maniera costruttiva gli strumenti base del calcolo probabilistico. Ad esempio, alla domanda “quante probabilità ho, se tiro 3 volte una moneta, di ottenere 3 volte testa?” la risposta “un ottavo“ è stata sostanzialmente immediata e concorde, e lo stesso si può dire per altri esercizi semplici proposti, riguardanti lanci di dadi e monete.
Il passaggio critico è stato semmai la formalizzazione di questi concetti, ed in particolare la definizione di un evento come un particolare sottoinsieme di un certo spazio Ω . In alcuni casi, sebbene i ragazzi riuscissero a risolvere senza troppi problemi un dato esercizio, non erano capaci di individuare con esattezza l'insieme Ω .

Per quel che riguarda invece il calcolo combinatorio, ci sono stati alcuni problemi più di fondo, dovuti principalmente alla scarsa abitudine a maneggiare oggetti come *n*-uple e prodotti di insiemi. Per darne un'idea, riporto alcune discussioni avvenute.

Fissato un insieme $A=\{a_{1,} a_2 \dots a_r\}$, e dopo che la docente ha spiegato cosa si intende per prodotto di insiemi ed in particolare per A^n , quando chiediamo a *quale insieme* appartiene la quaterna $(a_{1,} a_{2,} a_{3,} a_2)$, uno studente risponde immediatamente “ad A “. Gli diciamo che la risposta è sbagliata, ma lui commenta:
- “Se $a_{1,} a_{2,} a_3$ stanno in A , allora anche $(a_{1,} a_{2,} a_{3,} a_2)$ sta in A , no?!

Ovvio!”

Più avanti, quando chiediamo di stabilire una algoritmo per enumerare tutte le quaterne che si può creare con gli elementi di A, un altro ragazzo dice:
- Provo $(a_{\text{١}}, a_2, a_2, a_2)$, $(a_{\text{٢}}, a_3, a_3, a_3)$, $(a_{\text{٣}}, a_4, a_4, a_4)$ ... poi $(a_{\text{٢}}, a_1, a_{\text{١}}, a_1)$, $(a_{\text{٣}}, a_2, a_2, a_2)$ ... e poi mi sposto al secondo: $(a_{\text{١}}, a_2, a_3, a_3)$ ... e così via

Noi rispondiamo:
- Ma bisognerebbe trovare un metodo per essere sicuri di enumerarle *tutte*... magari una regola che dice “se hai appena trovato questa, la prossima è...”

E lui:
- “Ma un si po mi'a! Se per esempio cambio un simbolo a destra, che ne so di quello a sinistra?!”

L'*empasse* viene superata facendo notare al ragazzo che ad esempio per $A=\{0,1,2,3,4,5,6,7,8,9\}$ esiste un metodo, anche piuttosto semplice, per enumerare tutte le quaterne, ed è semplicemente contare da 0000 a 9999...

Segnalo che, nel momento in cui la docente presenta il coefficiente binomiale, un ragazzo dice “ah, sì, è la regola del *tre fattoriale*”. Sebbene parecchi altri ragazzi (anche di scuole diverse) abbiano confermato di avere studiato tale “regola” che, apparentemente, ha a che fare con i coefficienti binomiali, nessuno ci ha saputo spiegare in cosa consista esattamente e soprattutto da dove derivi il suo nome.

Terza fase: la dimostrazione del teorema.

Questa è, in sintesi, la dimostrazione del teorema che è stata fornita ai ragazzi:
1. Possiamo rappresentare ogni possibile configurazione dello spoglio dei voti come una *n*-upla contenente a volte A e b volte B (dove ovviamente $a+b=n$). Sia Ω l'insieme di tali *n*-uple. La sua cardinalità è data dalla formula del coefficiente binomiale ed è $\frac{n!}{a!(n-a)!}=\frac{n!}{a!b!}$.
2. Partizioniamo Ω in V e N , che contengono rispettivamente tutte le sequenze di scrutinio che vedono a sempre in vantaggio e tutte le restanti sequenze. Osserviamo che il nostro problema si riduce a trovare la probabilità che una sequenza scelta a caso sia in V .
3. Sia N_B l'insieme delle possibili sequenze di scrutinio in cui il primo voto va a B (osserviamo che $N_B \subset N$) e sia $N_A = N \setminus N_B$.
4. Osserviamo che se una *n*-upla corrisponde ad una sequenza di scrutinio appartenente a N_B , ci sarà una sua sottosequenza iniziale (o anche più di una, in tal caso consideriamo la più breve) che ha tante A quante B ; inoltre, sostituendo in tale sottosequenza ogni A con una B e viceversa otteniamo un'altra *n*-upla che corrisponde ad una sequenza di scrutinio appartenente a N_A . Questa è una relazione biunivoca tra N_A e N_B , che quindi hanno la stessa cardinalità.
5. Notiamo che è facile calcolare la dimensione di N_B , quindi quella di N_A , e di conseguenza quella di $\Omega \setminus (N_B \cup N_A) = V$.
6. Conoscendo le dimensioni di V e di Ω , ricaviamo la probabilità che

una sequenza di scrutinio casuale appartenga a V come $\frac{|V|}{|\Omega|}$.

Osservazione: la dimostrazione potrebbe essere semplificata eliminando il passo 5, evitando di calcolare la dimensione di N_B (e anche quella di Ω), limitandoci a notare che la probabilità che una *n*-upla qualsiasi appartenga a N_B è la probabilità che il primo voto scrutinato sia per B , ovvero $\frac{b}{n}$.

Quindi la probabilità che una *n*-upla appartenga a V è:

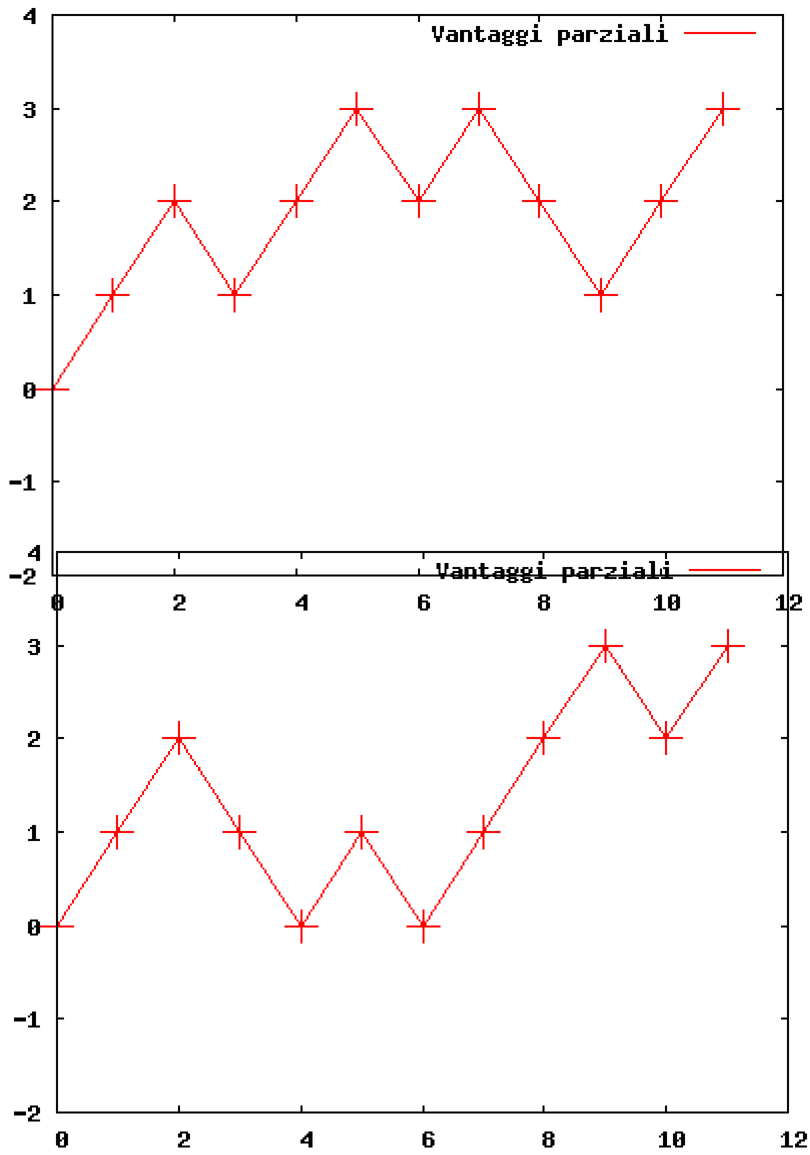
$$1-\frac{2b}{n}=\frac{a+b-2b}{n}=\frac{a-b}{a+b}$$.

Cionostante è stato scelto di evitare questo procedimento perché più astratto e per far sì che i ragazzi si esercitassero con i coefficienti binomiali.

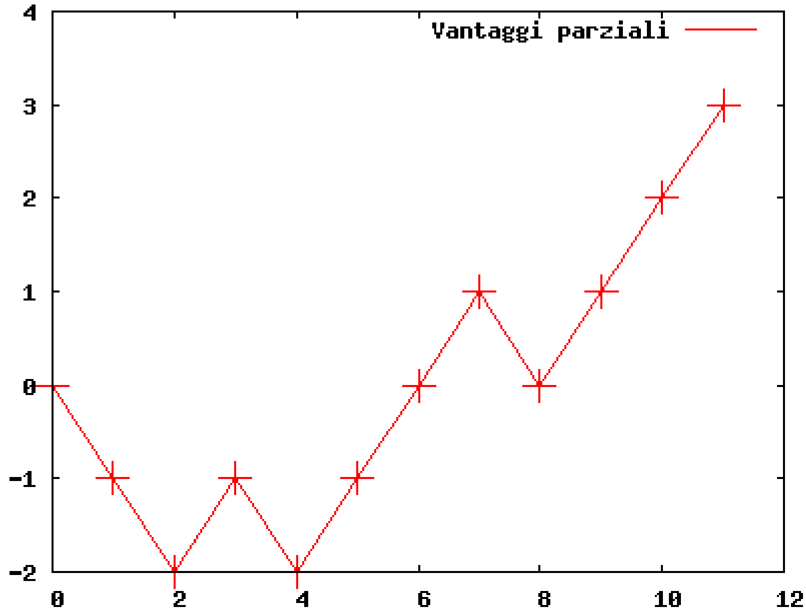
Passi 1,2,3: abbiamo incoraggiato i ragazzi a fare prove concrete con valori arbitrari di a e b , scrivendo le *n*-uple di A e B e caratterizzando quelle in cui il candidato A è sempre in vantaggio.

Passo 4: è stata molto utile, per facilitare la comprensione dei ragazzi, la rappresentazione grafica delle sequenze di scrutinio che abbiamo loro proposto, e che consisteva nel disegnare, in un piano cartesiano, una spezzata che assumesse in ogni $x \in \mathbb{N}, x \leqslant n$ il valore pari al vantaggio del candidato A dopo lo spoglio di x voti. Ad esempio, per $a=7, b=4$ possiamo rappresentare le seguenti successioni:

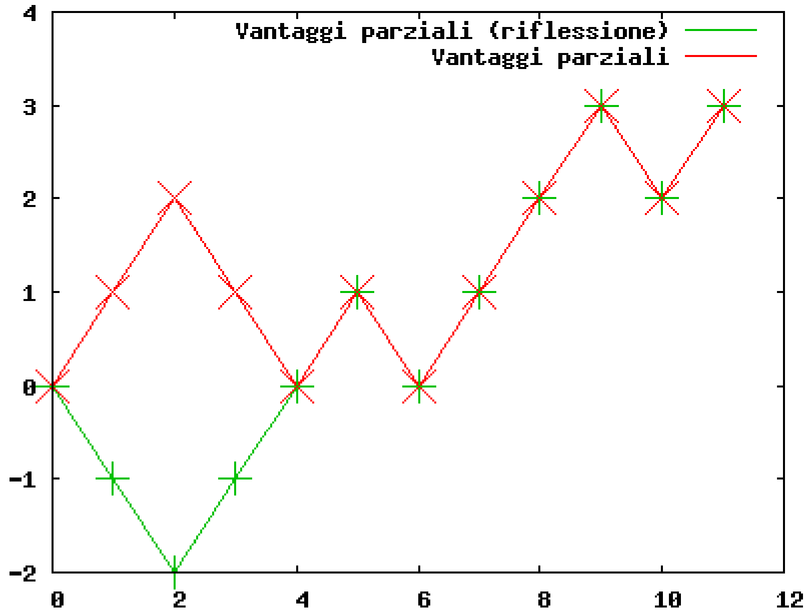
$(A\,A\,B\,A\,A\,B\,A\,B\,B\,A\,A)$



$(A\,A\,B\,B\,A\,B\,A\,A\,A\,B\,A)$



$(B\,B\,A\,B\,A\,A\,A\,B\,A\,A\,A)$



In questo modo risulta abbastanza evidente che la bigezione stabilita tra N_A e N_B è una sorta di *riflessione* (nell'immagine, la spezzata rossa rappresenta una sequenza di spoglio di N_A, e la spezzata verde la corrispondente di N_B).

Passo 5: N_B può essere visto come l'insieme delle sequenze di voto che sarebbero possibili se B avesse ottenuto solo $b-1$ voti (su $n-1$), ad ognuna delle quali viene aggiunto all'inizio un voto per B. La sua cardinalità è quindi (come per calcolare la dimensione di Ω si utilizza la formula del coefficiente binomiale):

$$\frac{a!}{(n-١)!((n-1)-a)!}=\frac{(b-١)!}{(n!)(n-(b-١))!}$$

Il resto sono solo calcoli, che sfruttano le proprietà del coefficiente binomiale:

$$|V|=|\Omega|-(|N_A|+|N_B|)=|\Omega|-٢|N_B|=\frac{n!}{(n-b)!b!}-٢\frac{(n-١)!}{((n-١)-(b-1))!(b-١)!}=$$

$$=\frac{n!-٢b(n-١)!}{a!b!}=\frac{(n-١)!(n-٢b)}{a!b!}=\frac{(n-1)!(a-b)}{a!b!}$$

Passo 6: Siamo in uno spazio con probabilità uniforme, quindi la probabilità che una sequenza qualsiasi di Ω si trovi in V è proprio:

$$\frac{|V|}{|\Omega|}=\frac{\frac{(n-1)!(a-b)}{a!b!}}{\frac{n!}{a!b!}}=\frac{a-b}{n}=\frac{a-b}{a+b}$$

Quarta fase: la sperimentazione al computer.

Il lavoro al computer è consistito sostanzialmente nel verificare che, sorteggiando in modo casuale un elevato numero di sequenze di scrutinio, il rapporto tra quelle in cui A è sempre in vantaggio e il totale di quelle sorteggiate si avvicina proprio a $\frac{a-b}{a+b}$.

Solo alcuni dei ragazzi avevano esperienza con linguaggi di programmazione, (principalmente il Pascal), così, per vincere anche il timore dei molti che per la prima volta si ritrovavano davanti ad un computer con sistema operativo Linux abbiamo utilizzato per il nostro esperimento dei semplici fogli di calcolo (Openoffice Calc), con cui più o meno tutti avevano una certa familiarità. Purtroppo questo ci ha costretti ad impostare il problema in un modo un po' meno lineare, ma è stata anche una buona occasione per conoscere alcune delle principali funzioni disponibili in un foglio di calcolo, le cui potenzialità sono solitamente poco conosciute anche a chi lo utilizza di frequente.

Queste sono quelle che abbiamo utilizzato:

if(condizione;oggetto1;oggetto2)

"**condizione**" è un test logico: se è verificato, la funzione restituisce **oggetto1**, altrimenti **oggetto2** (entrambi possono essere numeri, stringhe denotate da virgolette oppure altre funzioni).

rand()

Restituisce un valore pseudo-casuale compreso tra 0 e 1. Siccome noi avevamo bisogno di una sorta di estrazione con 2 sole possibilità, abbiamo associato ad un voto per A ogni valore di **rand()** inferiore a 0,5 e ad un voto per B ogni valore maggiore o uguale.

min(parametri)

Restituisce il valore minimo che trova all'interno delle celle **parametri**, che possono essere date come lista (ad es. **A1, B7, E8, C4**) oppure come blocco unico (ad es. **C7:C32** indica tutte le celle C7, C8, C9... C32).

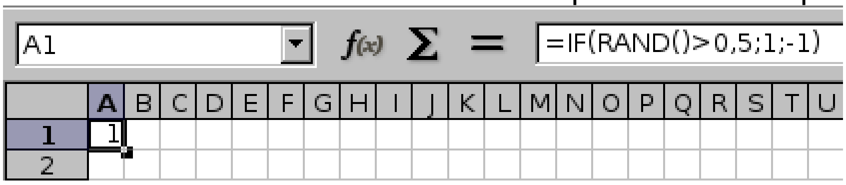
sum(parametri)

Restituisce la somma dei valori che si trovano all'interno delle celle **parametri** (le cui modalità di immissione sono le stesse che per **min**).

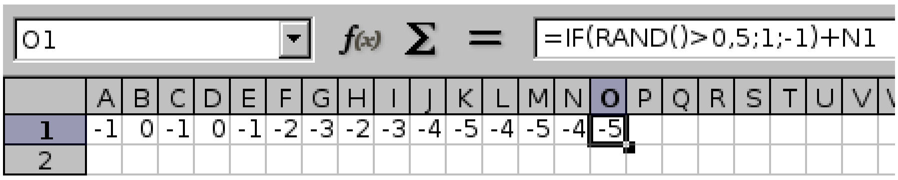
I ragazzi si sono distribuiti da soli oppure a gruppetti di due sulle postazioni dell'Aula 4. Su ogni postazione il foglio di calcolo è stato gestito in modo autonomo e quindi leggermente diverso, ma qui di seguito viene riassunta l'impostazione generale che si è voluto seguire.

Per facilità di calcolo, abbiamo deciso di rappresentare ogni voto per A con un +1 e ogni voto per B con un -1.

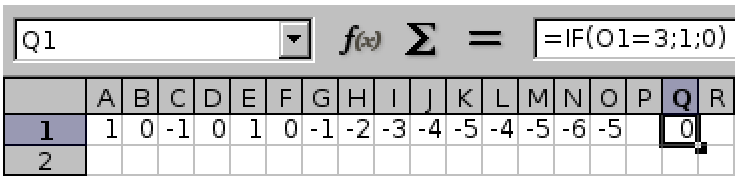
Questa è stata la formula che abbiamo immesso per estrarre il primo voto:



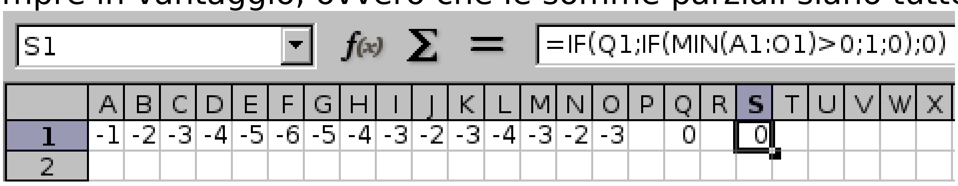
Nelle altre celle adiacenti abbiamo invece fatto in modo che comparisse direttamente la “somma parziale” dei primi voti, sommando un +1 o -1 al valore della cella a sinistra:



Quindi abbiamo scelto di prendere in esame un particolare valore di $a-b$ (su ogni postazione i ragazzi hanno scelto indipendentemente questo valore, tenendo conto del consiglio di non sceglierlo troppo alto). In questo caso il valore scelto è 3, e se la sequenza è ”interessante” (ovvero il vantaggio finale di A è proprio 3) la segnaliamo con un 1:



A questo punto in un'altra cella abbiamo scritto la formula per controllare che sia verificata la condizione su $a-b$ e contemporaneamente la condizione che A sia sempre in vantaggio, ovvero che le somme parziali siano tutte positive:



Abbiamo quindi ripetuto una riga siffatta un grande numero di volte (sfruttando il fatto che il foglio di calcolo considera i riferimenti alle celle, quando non specificato diversamente, relativi alla posizione della cella attuale e non assoluti):

S1000 | =IF(Q1000;IF(MIN(A1000:O1000)>0;1;0);0)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
991	1	0	1	2	3	2	3	4	5	4	3	4	5	6	5		0		0										
992	-1	0	-1	-2	-1	-2	-3	-4	-5	-4	-5	-6	-7	-6	-5		0		0										
993	1	0	-1	-2	-3	-2	-1	-2	-1	-2	-3	-2	-3	-4	-3		0		0										
994	1	0	-1	-2	-3	-4	-5	-4	-3	-2	-1	-2	-3	-2	-1		0		0										
995	1	0	-1	0	1	2	3	4	5	4	3	2	3	2	1		0		0										
996	-1	0	1	2	3	4	3	2	3	4	5	4	5	6	7		0		0										
997	1	2	1	0	-1	-2	-1	0	1	2	1	2	3	4	5		0		0										
998	-1	-2	-1	0	-1	0	1	0	1	0	-1	0	1	0	-1		0		0										
999	1	2	1	2	1	0	1	0	-1	0	-1	-2	-3	-4	-5		0		0										
1000	-1	-2	-1	-2	-3	-4	-3	-2	-3	-2	-3	-2	-3	-4	-3		0		0										
1001																													

E quindi abbiamo calcolato in una casella il numero delle volte che il vantaggio finale era quello desiderato, e in un'altra il numero di volte che il vantaggio era quello desiderato ed inoltre i vantaggi parziali erano tutti positivi:

S1002 | =SUM(S1:S1000)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
987	1	0	-1	-2	-3	-2	-3	-2	-3	-4	-3	-2	-1	-2	-3		0		0	
988	-1	0	1	2	3	2	3	4	5	4	3	2	3	2	1		0		0	
989	-1	-2	-3	-2	-3	-4	-5	-4	-5	-6	-7	-8	-9	-8	-9		0		0	
990	-1	-2	-3	-2	-3	-2	-3	-4	-5	-4	-5	-4	-5	-6	-7		0		0	
991	1	0	-1	-2	-1	-2	-1	-2	-3	-2	-3	-2	-1	-2	-3		0		0	
992	-1	0	-1	-2	-1	0	-1	0	1	2	3	4	3	4	5		0		0	
993	-1	-2	-3	-4	-5	-6	-5	-4	-5	-4	-5	-6	-7	-8	-7		0		0	
994	1	0	-1	-2	-1	-2	-1	0	1	0	-1	-2	-3	-4	-3		0		0	
995	-1	-2	-3	-4	-3	-4	-5	-6	-7	-6	-7	-8	-7	-6	-5		0		0	
996	1	0	1	0	1	2	1	0	-1	0	1	2	1	2	1		0		0	
997	1	2	1	0	-1	-2	-3	-2	-3	-4	-5	-4	-3	-4	-5		0		0	
998	1	0	1	0	-1	0	-1	0	1	0	1	2	1	2	1		0		0	
999	-1	0	-1	0	1	2	3	2	3	2	3	4	5	6	5		0		0	
1000	-1	-2	-3	-4	-3	-2	-1	0	-1	0	-1	-2	-1	-2	-1		0		0	
1001																				
1002																	141		27	
1003																				

Infine, in una cella a parte, abbiamo calcolato il rapporto tra queste due quantità:

997	-1	-2	-3	-2	-3	-2	-1	-2	-1	-2	-3	-4	-3	-2	-3		0		0		
998	1	2	1	2	1	0	1	0	-1	-2	-1	-2	-1	-2	-1		0		0		
999	1	2	1	2	3	2	3	2	3	2	1	0	1	2	1		0		0		
1000	1	2	1	0	-1	-2	-3	-2	-3	-2	-3	-4	-5	-4	-3		0		0		
1001																					
1002																	144		23		
1003																					
1004																					0,16
1005																					

Dalle immagini si può apprezzare il fatto che, ogni volta che viene fatta una modifica al foglio di calcolo, tutti i valori di **rand()** vengono ricalcolati; è stato perciò semplice fare diverse osservazioni e appurare che il valore di questa ultima cella si aggirasse proprio intorno a $\frac{a-b}{a+b}$ (nel caso osservato, $3/15=0{,}2$).

Pietro Battiston